Paper ML05

Machine Learning Algorithms to Explore the Relationship Between Primary Obsessive–Compulsive Disorder (OCD) and Co-morbid Depression Symptoms

Ajay Kumar Singh, ICON India Clinical Research Ltd, Bangalore, India
Sabyasachi Mukherjee, ICON India Clinical Research Ltd, Bangalore, India

ABSTRACT

Obsessive–Compulsive Disorder (OCD) and Major Depressive Disorder (MDD) frequently co-occur, with comorbidity estimates exceeding 40%. This comorbidity contributes to worse symptom severity, functional impairment, and treatment resistance. The present study applied machine learning (ML) and data mining techniques to examine predictive relationships between depressive symptoms and obsessive, compulsive, and overall OCD severity in adults with a primary diagnosis of OCD. Secondary data from 408 patients receiving treatment at a specialty OCD program were analyzed using average obsession, compulsion, and OCD scores as outcome variables. Predictive modeling was performed using Artificial Neural Networks (ANN), Support Vector Machines (SVM), and Random Forests (RF). Sensitivity analysis was used to identify key depressive features contributing to OCD symptom dimensions. SVM demonstrated superior predictive accuracy across all outcome variables, outperforming ANN in terms of lower mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE). Sadness, sleep latency, decreased appetite, and increased appetite were most important for predicting obsession, whereas psychomotor agitation, psychomotor retardation, anhedonia, and fatigue were dominant predictors of compulsion. When obsession and compulsion were combined, appetite disturbance, suicidality, and anhedonia emerged as prominent contributors. These findings highlight transdiagnostic affective, neurovegetative, and psychomotor processes linking depressive features with OCD symptom severity, underscoring the importance of integrated assessment and treatment.

1. INTRODUCTION

Obsessive-compulsive disorder is recognized by a variety of symptoms which is characterized by repetitive intrusive thoughts and urges, and by compulsive thoughts and actions designed to relieve the distress caused by these obsessions (American Psychiatric Association, 2013). These recurring obsessions or compulsions bring the person in great distress. However, it becomes a disorder when it interferes with the person's normal routine, occupational functioning, usual social activities or relationships. A patient with OCD may have an obsession, a compulsion, or both. Among American adults, OCD has 12-month and lifetime prevalence rates of 1.2% and 2.3%, respectively (Ruscio et al. 2010). Several studies show that among youth diagnosed with Obsessive-Compulsive Disorder (OCD), the rates of co-occurring depressive disorders or depressive symptoms are substantial. For example, in a clinical sample of children/adolescents with OCD, depressive disorders ranged from 13 %-46% depending on measurement and sample (Canavera, K., et al. (2008), Storch, E. A., et al. (2012)).

Depression, defined by persistent low mood, anhedonia, cognitive impairment, and neurovegetative symptoms, is one of the most prevalent and disabling global mental health conditions. Epidemiological studies indicate that approximately 2–3% of the world's population experiences OCD, whereas more than 300 million individuals suffer from depression worldwide (APA, 2022).

Comorbidity between OCD and depression is substantial. Over 40% of individuals with OCD also meet criteria for depressive disorder, and this overlap contributes to higher functional impairment, elevated healthcare utilization, prolonged illness duration, and increased suicide risk (Ruscio et al., 2010). Despite this strong association, the specific symptom-level relationships between OCD dimensions and depressive features remain underexplored.

Machine learning (ML) and data mining techniques are increasingly applied in the field of mental health to improve the diagnosis, treatment, and monitoring of conditions like obsessive-compulsive disorder (OCD) and depression. Machine learning (ML) offers a powerful framework for understanding these symptom interdependencies, particularly the nonlinear relationships that conventional statistics may overlook. ML algorithms can help to detect OCD and depression earlier than traditional clinical methods by analyzing patterns through early detection and diagnosis (Jiyeong Kim et., al 2025, Tubío-Fungueiriño M 2022).

The present study investigates predictive relationships between depression symptoms (measured via QIDS-SR) and obsession, compulsion, and combined OCD severity (measured via Y-BOCS-SR) using multiple ML algorithms. Additionally, sensitivity analysis provides insights into which depressive symptoms most strongly influence OCD dimensions.

1.1. LITERATURE REVIEW

OCD and depression share cognitive, affective, and neurobiological features. Obsessions are often associated with guilt, rumination, intrusive thoughts, and impaired cognitive control—processes, which are also central to depression. Compulsions, though behavioral, are reinforced by emotional dysregulation, anxiety reduction mechanisms, and maladaptive avoidance patterns. Neurobiologically, abnormalities in cortico-striato-thalamo-cortical (CSTC) circuits are implicated in OCD, while depressive disorders involve prefrontal–limbic dysregulation.

Prior research has shown:

- Depressive symptoms exacerbate OCD severity and treatment resistance (Abramowitz et al., 2009).
- Sleep disturbance, psychomotor changes, and anhedonia influence obsessive–compulsive symptom expression.
- Symptom-level (rather than diagnosis-level) associations may offer better predictive value for clinical outcomes.

Machine learning has gained traction in psychiatry for uncovering such symptom-level interactions. ANN, SVM, and RF models are particularly useful for their ability to model nonlinear relationships and identify complex variable interactions.

2. MATERIAL AND METHODOLOGY

2.1. MATERIAL

The present study is based on the secondary data on 408 adults beginning treatment for OCD at Rogers Memorial Hospital having 26 factors or variables. The dataset “Rogers” is available for download from the "MPsychoR" library package, developed under R statistical software contains several psychological data sets with descriptions. The descriptions of the variables are provided in Tables 1 and 2. There are 10 factors measuring obsession and compulsion behavior and 16 factors assessing depressive parameters within the patients in the current data. The Children's Yale-Brown Obsessive Compulsive Scale (CY-BOCS; Scahill et al. 1997) and the Quick Inventory of Depressive Symptomatology (QIDS-SR; Rush et al. 2003) are two self-report questionnaires used to evaluate the OCD and depression parameters among different individuals.

TABLE 1 DESCRIPTION OF VARIABLES MEASURING OBSESSION AND COMPULSION RESPECTIVELY

VARIABLE NAME	ANALYSIS VARIABLE	DESCRIPTION
Obtime	obtime	Time occupied by obsessive thoughts
Obinterfer	obinterfer	Interference in one’s life due to obsessive thoughts
Obdistress	obdistress	Distress associated with obsessive thoughts
Obresist	obresist	Difficulty resisting obsessions
Obcontrol	obcontrol	Difficulty controlling obsessions
Comptime	comptime	Time spent performing compulsive behaviors
Compinterf	compinterf	Interference in one’s life due to compulsive behaviors
Compdis	compdis	Distress associated with compulsive behaviors
Compresis	compresis	Difficulty resisting compulsions
Compcont	compcont	Difficulty controlling compulsions

TABLE 2 DESCRIPTION OF VARIABLES MEASURING DEPRESSION

VARIABLE NAME	ANALYSIS VARIABLE	DESCRIPTION
Onset	Sl_on	Difficulty falling asleep
Middle	SL_mid	Sleep during the night
Late	SL_la	Waking up too early
Hypersom	Hypersom	Sleeping too much

Sad	Sad	Feeling sad
Decappetite	AP_De	Decreased appetite
Incappetite	AP_In	Increased appetite
Weightloss	We_loss	Decreased weight within the last two weeks
Weightgain	We_gain	Increased weight within the last two weeks
Concen	Con	Impairment in concentration/decision-making
Guilt	Guilt	Self-blaming view of oneself
Suicide	Suicide	Suicidal thoughts, plans or attempts
Anhedonia	Anhedonia	Anhedonia
Fatigue	Fatigue	Fatigue
Retard	Retard	Psychomotor retardation
Agitation	Agitation	Agitation

2.2. METHODOLOGY

The principal motivation of this presentation is twofold: one is to explore different predictive functional relationships between (i) Obsession and depression parameters (ii) Compulsion and depression parameters (iii) together OCD and depression; another one is to develop the understanding and implication of ML and data mining in psychopathology.

- Take an average score for each patient obtained from the questionnaire measuring OCD symptoms to get an average/mean unbiased score of obsession, compulsion and OCD.
- These average scores for obsession, compulsion and OCD were used as the response/target variable to find the best predictive non-linear relationship (using ML and data mining) between the target variable (obsession, compulsion and OCD) and input variables (depression parameters).
- For important feature selection, sensitivity analysis has been used. Sensitivity analysis is a technique that is used to determine how the values of independent variables value will determine the value of a particular dependent variable under a given set of assumptions.

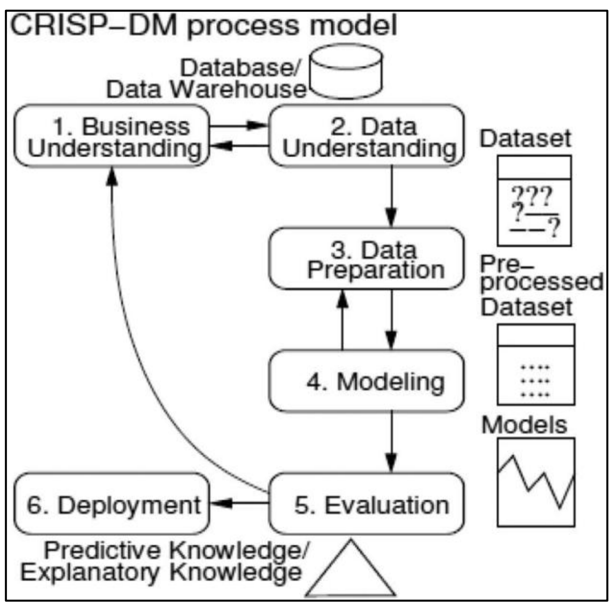


Figure 1. The CRISP-DM tool

Data mining (DM) is an iterative process consisting of six phases described in CRISP-DM (Chapman et. al. 2000), a tool-neutral methodology that has been approved by the data analytics industry (e.g., SPSS, DaimlerChryslyer). According to this, any DM project has been partitioned into the following six phases: (1) Business understanding (2) Data understanding (3) Data preparation (4) Modelling (5) Evaluation and (6) Deployment. These stages are also shown in Figure 1.

To complete a specific regression task, this study focuses on modelling and evaluation steps by utilizing DM techniques with an emphasis on various ML algorithms viz. Artificial neural network (ANN), Support vector machine (SVM), Random Forest (RF).

2.2.1 ARTIFICIAL NEURAL NETWORK (ANN)

ANN means the popular Multilayer Perceptron (MLP). MLPs (Figure 2) has proven to be an effective way to solve classification and regression tasks. A major concern in their use is the difficulty in defining the proper network for a specific application, due to the sensitivity of the initial conditions and to over-fitting and under-fitting problems that limit their generalization capability. Moreover, time and hardware constraints may seriously reduce the degrees of freedom in the search for a single optimal network. A very promising way to partially overcome such drawbacks is the use of MLP ensembles (MLPE), averaging and voting techniques are largely used in classical statistical pattern recognition and can be fruitfully applied to MLP classifiers. For classification and regression problems, MLPE is used, which is a combination of MLP models, and it is observed that in many situations, MLPE gives better results than any of its single MLPs.

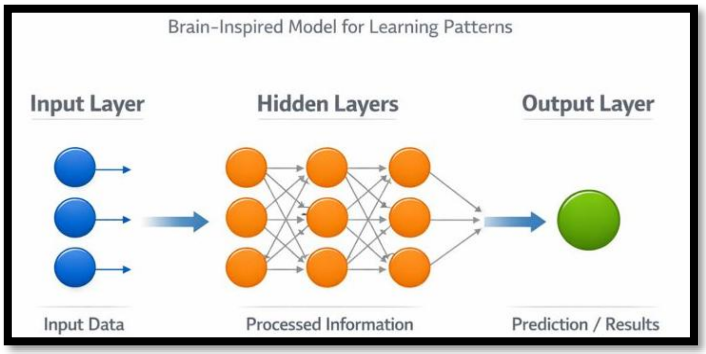


Figure 2. Artificial Neural Network

2.2.2 SUPPORT VECTOR MACHINE (SVM)

SVM has theoretical advantages over NNs, such as the absence of local minima in the learning phase (Hastie, Tibshirani, and Friedman 2008). The basic idea is to transform the input $x \in \Re^{I}$ into a high m-dimensional feature space by using a nonlinear mapping. Then, the SVM finds the best linear separating hyperplane, related to a set of support vector points, in the feature space (Figure- 3).

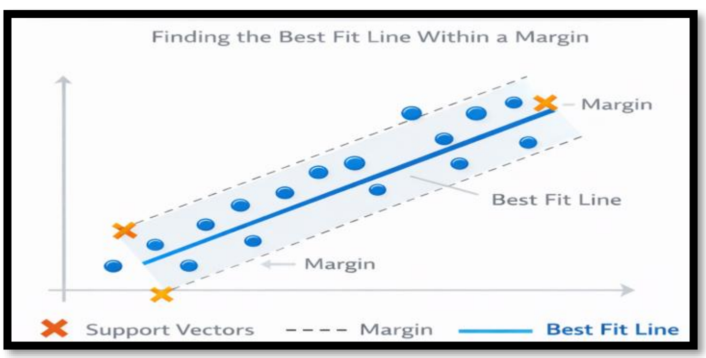


Figure 3. Support Vector Machine

2.2.3 RANDOM FOREST (RF)

Random Forest is an ensemble machine-learning algorithm that builds multiple decision trees using random subsets of data and features, then combines their predictions to produce a more accurate and robust result. For classification, it uses majority voting; for regression, it averages the outputs. By introducing randomness and combining many trees, Random Forest reduces overfitting and improves generalization compared to a single decision tree (Figure 4).

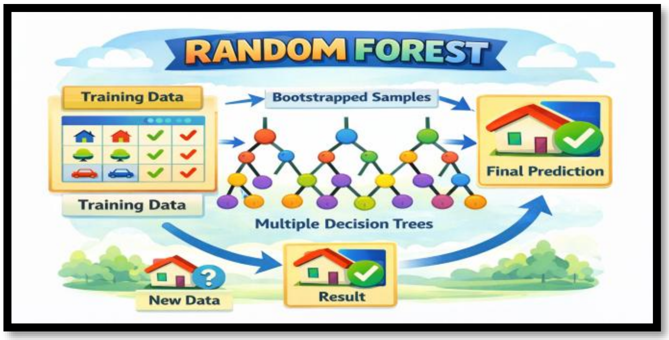


Figure 4. Random Forest

2.2.4 FEATURE SELECTION USING SENSITIVITY ANALYSIS

The sensitivity analysis (Figure 5) is a simple procedure that is applied after the training procedure and analyses the model responses when a given input is changed. Let $y_{a,j}$ denote the output obtained by holding all input variables at their average values except x_a, which varies through its entire range ($x_{a,j}$, with $j \in \{1,2,\ldots L\}$ levels).The variance (V_a) of $y_{a,j}$ used as a measure of input relevance (Kewley, Embrechts, and Breneman 2000). For classification task, if N_c>2 (multi-class), it sets as the sum of the variances for each output class probability ($p(c)_{a,j}$) . A high variance (V_a) suggests a high x_a relevance, thus the input relative importance (R_a) is given by:

$$R_a = \frac{V_a}{\sum_{i=1}^{I} V_i \times 100(\%)}$$

For a more detailed analysis, Variable Importance Chart and Variable Effect Characteristic (VEC) curve has been proposed by Cortez et. al. 2009, which plots importance of all variables in terms of R_a for the corresponding response variable and the $x_{a,j}$ values (x-axis) versus the $y_{a,j}$ predictions (y-axis) respectively.

How it works:

- Each feature is varied slightly while keeping others constant.
- The change in model output is measured.
- Features causing larger output changes are considered more important.

Why it's useful:

- Identifies key features that strongly impact predictions
- Improves model interpretability
- Helps reduce irrelevant or redundant features

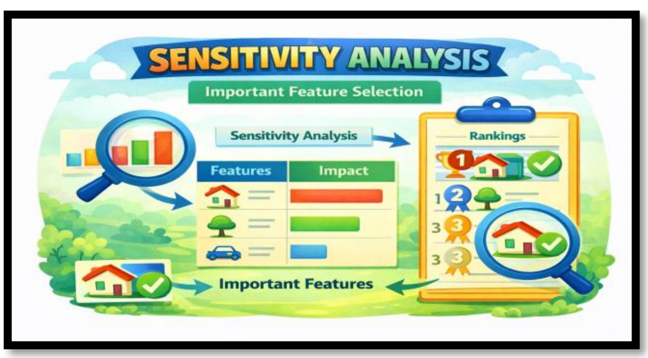


Figure 5. Sensitivity Analysis

2.2.5 EVALUATION METRICS FOR MODEL PREDICTION

In this presentation, five metrics were utilized to compare the effectiveness of the DM methods, and they are sum of absolute error (SAE), mean absolute error (MAE), mean square error (MSE), sum of square error (SSE), and root mean square error (RMSE) respectively.

TABLE 3 PERFORMANCE EVALUATION METRICS

Metric	Full Form	Formula	Description
SAE	Sum of Absolute Errors	$\sum \lvert y_i - \hat{y}_i \rvert$	Total absolute error
MAE	Mean Absolute Error	$\frac{1}{n} \sum \lvert y_i - \hat{y}_i \rvert$	Average absolute error
MSE	Mean Squared Error	$\frac{1}{n} \sum (y_i - \hat{y}_i)^2$	Average squared error
SSE	Sum of Squared Errors	$\sum (y_i - \hat{y}_i)^2$	Total squared error
RMSE	Root Mean Squared Error	$\sqrt{\frac{1}{n} \sum (y_i - \hat{y}_i)^2}$	Root of MSE

3. RESULTS AND DISCUSSION

3.1. OBSESSION: RESULTS AND DISCUSSION

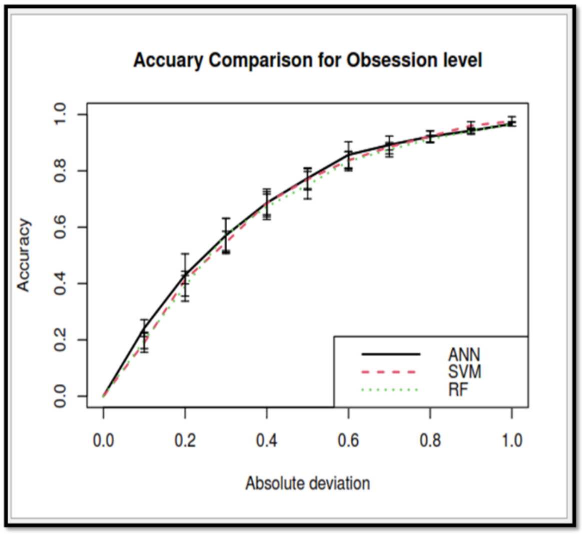
3.1.1 RESULTS

TABLE 4 Neural Network (NN)

RUN	SAE	MAE	MSE	SSE	RMSE
1	44.58697	0.32784	0.20141	27.39303	0.44879
2	47.86138	0.35192	0.21202	28.83574	0.46046
3	45.43233	0.33406	0.20509	27.89334	0.45287
4	41.10860	0.30226	0.18808	25.57892	0.43368
5	47.52364	0.34943	0.21078	28.66735	0.45911
MEAN	45.30258	0.33310	0.20348	27.67368	0.45098
95% CI	± 3.37763	± 0.02483	± 0.01195	± 1.62520	± 0.01336

TABLE 5 Support Vector Machine (SVM)

RUN	SAE	MAE	MSE	SSE	RMSE
1	43.78497	0.32194	0.18769	25.52613	0.43323
2	45.14387	0.33194	0.17889	24.32959	0.42295
3	47.25219	0.34744	0.19642	26.71385	0.44319
4	46.79813	0.34410	0.18656	25.37252	0.43192
5	45.74196	0.33633	0.19055	25.91608	0.43653
MEAN	45.74423	0.33635	0.18802	25.57163	0.43357
95% CI	± 1.70969	± 0.01257	± 0.00791	± 1.07668	± 0.00914



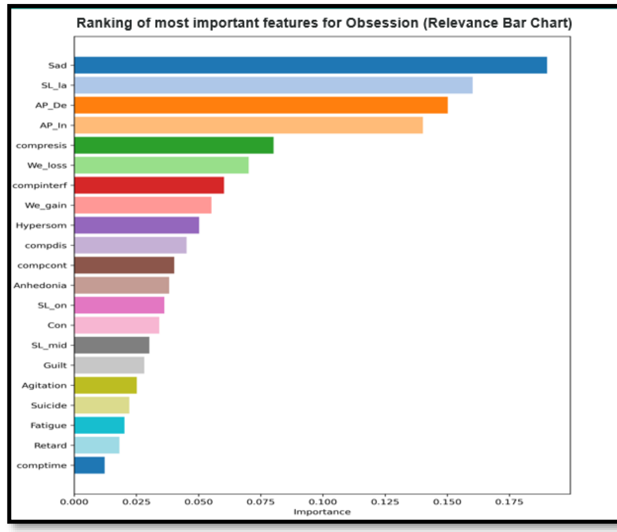


Figure 6: Obsession: (a) Accuracy Comparison (for best three models) (b) Feature Selection using Sensitivity Analysis (under best model SVM)

3.1.2 DISCUSSION

- SVM model is better for overall accuracy (lower MSE, SSE, RMSE) with Stability advantages of SVM over ANN (narrower confidence intervals)
- Obsessive symptoms showed strongest associations with affective and physiological factors, especially depressed mood, sleep latency, and appetite disturbances.
- Sadness emerged as the most influential predictor, underscoring the close link between obsessions and depressive affect.
- Findings align with Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5), American Psychiatric Association (APA) models of Obsessive–Compulsive Disorder (OCD) and Major Depressive Disorder (MDD), which frequently co-occur.
- The strong obsession–sadness link supports evidence that comorbid depression increases obsessional severity and functional impairment.
- Sleep latency was highly associated with obsession, suggesting intrusive thoughts and cognitive hyperarousal disrupt sleep initiation.
- Sleep disturbance, common in both OCD and depression, may further impair emotional regulation and reinforce obsessive thinking.
- Both increased and decreased appetite were linked to obsession, reflecting heterogeneous stress and coping responses.
- Appetite disturbance may indicate neurovegetative dysregulation or maladaptive coping (e.g., emotional eating).
- Overall, mood and neurovegetative symptoms outweighed compulsive behaviors, supporting a transdiagnostic framework.
- Clinically, targeting mood and sleep disturbances may indirectly reduce obsessive symptom severity.

3.2. COMPULSION: RESULTS AND DISCUSSION

3.2.1 RESULTS

TABLE 6 Neural Network (NN)

RUN	SAE	MAE	MSE	SSE	RMSE
1	51.46456	0.37841	0.23414	31.84389	0.48388
2	53.84281	0.39590	0.26657	36.25469	0.51631
3	52.69002	0.38742	0.28896	39.29864	0.53755

4	47.93030	0.35242	0.25808	35.09957	0.50802
5	53.66637	0.39460	0.30172	41.03443	0.54929
MEAN	51.91881	0.38175	0.26989	36.70624	0.51901
95% CI	± 3.00736	± 0.02211	± 0.03288	± 4.47178	± 0.03179

TABLE 7 Support Vector Machine (SVM)

RUN	SAE	MAE	MSE	SSE	RMSE
1	49.49664	0.36394	0.22367	30.41930	0.47293
2	44.61625	0.32806	0.19946	27.12666	0.44661
3	52.52254	0.38619	0.29096	39.57140	0.53941
4	48.63455	0.35760	0.27606	37.54477	0.52541
5	47.10926	0.34639	0.24459	33.26480	0.49456
MEAN	48.47585	0.35644	0.24695	33.58539	0.49578
95% CI	± 3.63104	± 0.02669	± 0.04640	± 6.31149	± 0.04696

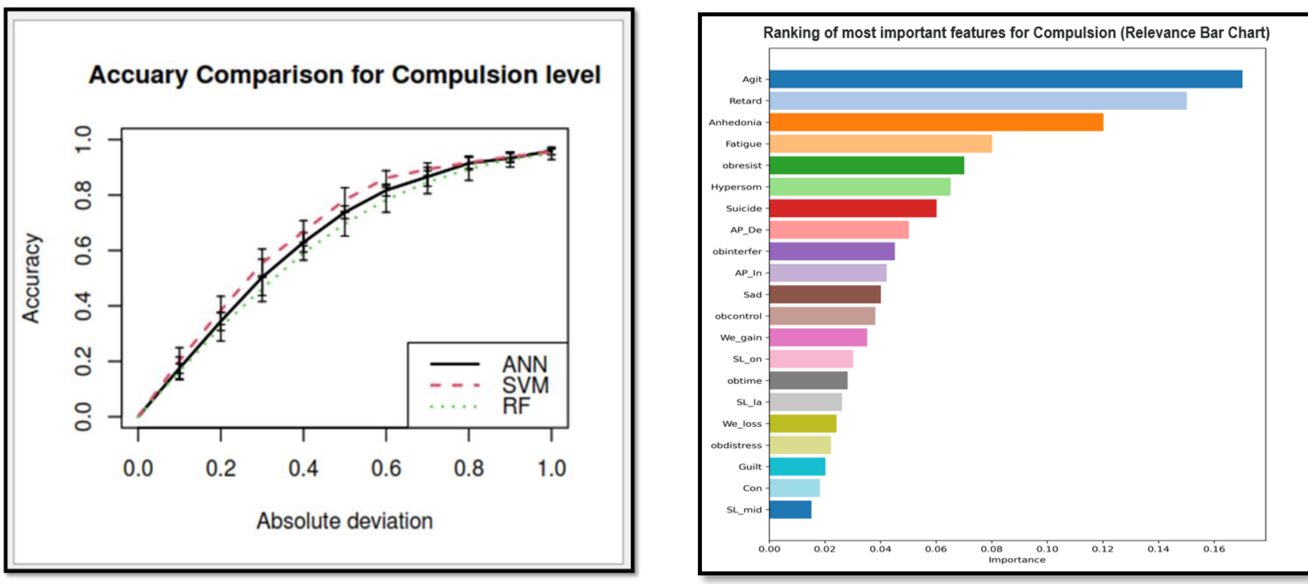


Figure 7: Compulsion: (a) Accuracy Comparison (for best three models) (b) Feature Selection using Sensitivity Analysis (under best model SVM)

3.2.2 DISCUSSION

- SVM achieves lower Mean SAE, MAE, MSE, SSE, and RMSE compared to NN, indicating superior predictive performance.
- SVM demonstrates better average accuracy across most metrics. Therefore, SVM outperforms NN for the Compulsion dataset.
- Compulsive symptoms were most strongly associated with psychomotor agitation, psychomotor retardation, anhedonia, and fatigue.
- This pattern suggests compulsions are closely linked to motor activity and energy dysregulation, rather than affective distress alone.

- Psychomotor agitation emerged as the strongest predictor, consistent with DSM-5 descriptions of compulsions as driven, repetitive behaviors performed to relieve internal tension or anxiety.
- The association with psychomotor retardation indicates heterogeneity in compulsive presentation, with behaviors becoming rigid, effortful, and automatic, overlapping with depressive motor slowing.
- Anhedonia was strongly related to compulsion, supporting the role of negative reinforcement rather than pleasure in maintaining compulsive behaviors.
- This aligns with DSM-5 criteria emphasizing that compulsions are not experienced as enjoyable but are performed to reduce distress or prevent feared outcomes.
- Fatigue reflects the cognitive and physical burden of repetitive rituals and reduced capacity to resist compulsive urges.
- Overall, findings support a DSM-5–based view of compulsions as repetitive behaviors embedded in psychomotor and neurovegetative dysfunction.
- Clinically, assessment and treatment should address motor and energy-related disturbances, especially in the presence of comorbid depressive features.

3.3. OCD: RESULTS AND DISCUSSION

3.3.1 RESULTS

TABLE 8 Neural Network (NN)

RUN	SAE	MAE	MSE	SSE	RMSE
1	56.64063	0.41648	0.29376	39.95120	0.54200
2	54.41534	0.40011	0.25276	34.37592	0.50276
3	51.15987	0.37618	0.21654	29.45002	0.46534
4	63.41354	0.46628	0.34535	46.96785	0.58767
5	62.69511	0.46099	0.33258	45.23145	0.57670
MEAN	57.66490	0.42401	0.28820	39.19529	0.53489
95% CI	± 6.57830	± 0.04837	± 0.06703	± 9.11621	± 0.06347

TABLE 9 Support Vector Machine (SVM)

RUN	SAE	MAE	MSE	SSE	RMSE
1	54.22839	0.39874	0.27900	37.94331	0.52820
2	53.87476	0.39614	0.23949	32.57030	0.48937
3	53.60938	0.39419	0.25247	34.33593	0.50246
4	56.96495	0.41886	0.27753	37.74375	0.52681
5	53.65952	0.39456	0.25063	34.08609	0.50063
MEAN	54.46740	0.40050	0.25982	35.33588	0.50950
95% CI	± 1.75981	± 0.01294	± 0.02180	± 2.96488	± 0.02135

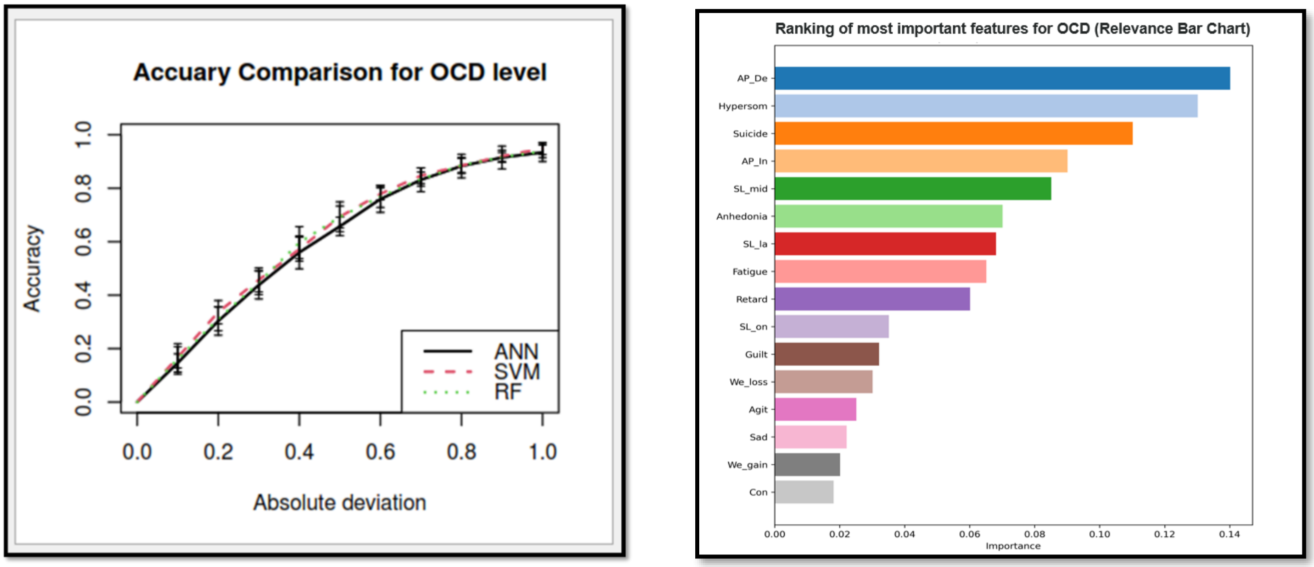


Figure 8: OCD: (a) Accuracy Comparison (for best three models) (b) Feature Selection using Sensitivity Analysis (under best model SVM)

3.3.2 DISCUSSION

- SVM achieves lower Mean SAE, MAE, MSE, SSE, and RMSE compared to NN
- NN exhibits higher variability, reflected by wider confidence intervals
- SVM shows more stable and accurate performance for OCD prediction. Overall, SVM outperforms NN across all evaluated error metrics.
- Decreased appetite (AP_De) showed the highest importance, indicating strong influence on OCD symptom severity.
- Hypersomnia emerged as the second most influential factor, highlighting significant sleep-related dysregulation.
- Suicidality ranked third, suggesting severe depressive distress is closely linked to obsessive–compulsive symptoms.
- Increased appetite (AP_In) was also highly influential, reflecting neurovegetative appetite disturbances.
- Middle insomnia (SL_mid) followed closely, reinforcing the role of sleep disruption in OCD.

Overall interpretation: The most influential predictors are predominantly depressive and neurovegetative symptoms, particularly appetite disturbance, sleep dysregulation, and suicidality, underscoring that obsession–compulsion symptoms are strongly embedded within depressive symptom networks rather than isolated anxiety processes.

4. CONCLUSION

This study demonstrates that:

- ML techniques, particularly SVM, effectively model complex relationships between depressive and OCD symptoms.
- Different depressive symptoms predict obsession and compulsion uniquely. For obsession, sadness is relevant when it is persistent
- Affective (sadness), neurovegetative (sleep and appetite), and psychomotor (agitation/retardation) symptoms drive OCD severity.
- Low error & variance for obsession compared to compulsion is showing that symptoms for obsession have better association with depressive symptoms
- Comorbid depression shapes OCD presentation in systematic, predictable ways.

Clinically, the results support the adoption of transdiagnostic treatment approaches addressing mood, sleep, and psychomotor functioning alongside traditional OCD interventions such as ERP (Exposure and Response Prevention).

5. REFERENCES

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th ed.). *American Psychiatric Publishing.*

Canavera, K. E., Ollendick, T. H., Ehrenreich May, J. T., & Pincus, D. B. (2010). Clinical correlates of comorbid obsessive–compulsive disorder and depression in youth. *Child Psychiatry & Human Development*, *41*(6), 583-594.

Ruscio, A. M., Stein, D. J., Chiu, W. T., & Kessler, R. C. (2010). The epidemiology of obsessive–compulsive disorder in the National Comorbidity Survey Replication. *Molecular Psychiatry, 15*(1), 53–63.

Storch, E. A., Larson, M. J., Ehrenreich-May, J., Arnold, E. B., Jones, A. M., Renno, P., ... & Wood, J. J. (2012). Peer victimization in youth with autism spectrum disorders and co-occurring anxiety: relations with psychopathology and loneliness. *Journal of Developmental and Physical Disabilities*, *24*(6), 575-590.

American Psychiatric Association. (2022). Diagnostic and statistical manual of mental disorders (5th ed., text rev.; DSM-5-TR). *American Psychiatric Publishing.*

Kim, J., Pacheco, J. P. G., Golden, A., Aboujaoude, E., van Roessel, P., Gandhi, A., … & Stirman, S. W. (2025). *Artificial intelligence in obsessive-compulsive disorder: A systematic review*. *Current Treatment Options in Psychiatry, 12*(23). https://doi.org/10.1007/s40501-025-00359-8

Tubío-Fungueiriño, M., Cernadas, E., Gonçalves, O. F., Segalàs, C., Bertolín, S., Mar-Barrutia, L., Real, E., Fernández-Delgado, M., Menchón, J. M., Carvalho, S., Alonso, P., Carracedo, A., & Fernández-Prieto, M. (2022). *Viability study of machine learning-based prediction of COVID-19 pandemic impact in obsessive-compulsive disorder patients*. *Frontiers in Neuroinformatics, 16*, Article 807584. https://doi.org/10.3389/fninf.2022.807584

Abramowitz, J. S., Taylor, S., & McKay, D. (2009). Obsessive–compulsive disorder. The Lancet, 374(9688), 491–499. https://doi.org/10.1016/S0140-6736(09)60240-3

Rush, A. J., Trivedi, M. H., Ibrahim, H. M., Carmody, T. J., Arnow, B., Klein, D. N., Markowitz, J. C., Ninan, P. T., Kornstein, S., Manber, R., Thase, M. E., Kocsis, J. H., & Keller, M. B. (2003). The 16-item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): A psychometric evaluation in patients with chronic major depression. *Biological Psychiatry, 54*(5), 573–583. https://doi.org/10.1016/S0006-3223(02)01866-8

Scahill, L., Riddle, M. A., McSwiggin-Hardin, M., Ort, S. I., King, R. A., Goodman, W. K., Cicchetti, D., & Leckman, J. F. (1997). Children's Yale-Brown Obsessive Compulsive Scale: Reliability and validity. *Journal of the American Academy of Child & Adolescent Psychiatry, 36*(6), 844–852. *https://doi.org/10.1097/00004583-199706000-00023*
Chapman, P., J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, and R. Wirth. 2000. CRISP-DM 1.0: Step-by-step data mining guide. CRISP-DM consortium.

T. Hastie, R. Tibshirani, and J. Friedman. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer-Verlag, NY, USA, 2nd ed., 2008.

Kewley, R. H., Embrechts, M. J., & Breneman, C. (2000). Data strip mining for the virtual design of pharmaceuticals with neural networks. *IEEE Transactions on Neural Networks*, *11*(3), 668-679.

Cortez, P., Cerdeira, A., Almeida, F., Matos, T., & Reis, J. (2009). Modeling wine preferences by data mining from physicochemical properties. *Decision support systems*, *47*(4), 547-553.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Author Name: Ajay Kumar Singh
Company: ICON PLC
Address: 3rd Floor, PRESTIGE BLUE CHIP SOFTWARE PARK, Block 1, Marigowda Rd, Dairy Colony, Adugodi, Bengaluru, Karnataka 560030, India
Work Phone: NA
Email: Ajaykumar.singh@iconplc.com
Website: https://www.iconplc.com/

Co-Author Name: Sabyasachi Mukherjee
Company: ICON PLC
Address: 3rd Floor, PRESTIGE BLUE CHIP SOFTWARE PARK, Block 1, Marigowda Rd, Dairy Colony, Adugodi, Bengaluru, Karnataka 560030, India
Work Phone: NA
Email: Sabyasachi.Mukherjee@iconplc.com
Website: https://www.iconplc.com/